The logo for Archive-IT, featuring a large white letter 'A' with a blue circle in the center, set against a light blue square background.

**Starting shortly** ⌚

**ARCHIVE-IT**



The logo for Archive-IT is centered on the page. It consists of a light blue square background. Inside the square is a large, white, stylized letter 'A'. A small, solid blue circle is positioned in the center of the 'A'. Overlaid on the center of the 'A' is the text 'Quality Assurance' in a bold, black, sans-serif font. Below the 'A' and the text, the words 'ARCHIVE-IT' are written in a white, bold, sans-serif font.

**Quality Assurance**

**ARCHIVE-IT**

# TODAY'S PLAN

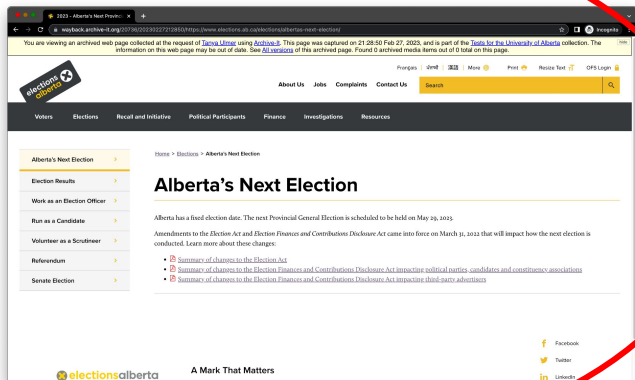
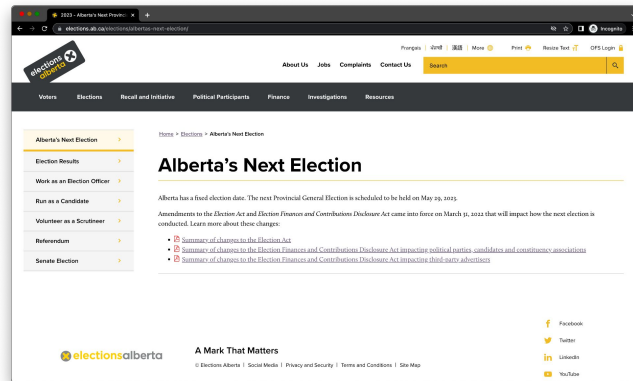


Image courtesy of [Volodymyr Hryshchenko](#) on Unsplash

## Quality Assurance

- Review Crawl Reports
- Browse Wayback pages
- Patch Crawls
- Helper Seeds
- Questions

# Web Archiving



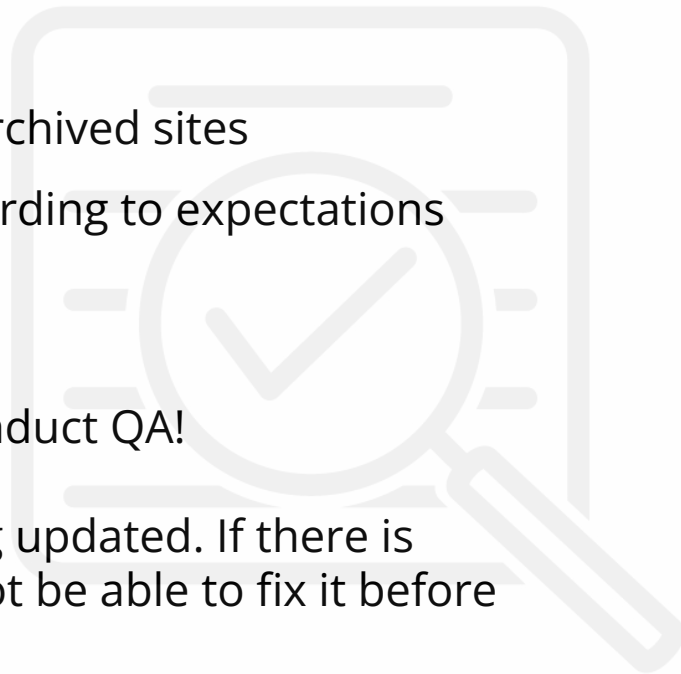
## Why Quality Assurance (QA)?

The web can sometimes be challenging to archive.  
Quality assurance is important for:

- ❖ checking for completeness and quality of archived sites
- ❖ checking that the archived sites replay according to expectations
- ❖ improving the quality of your capture

If possible, don't wait too long to review and conduct QA!

Live websites are constantly changing and being updated. If there is something missing from the crawl, you might not be able to fix it before the content changes or is removed.



**Crawl Report:** A report generated with detailed information about what's inside your crawl.

**Host:** Where web content is stored, designated by its Internet host name. E.g. <https://archive-it.org/> (host bolded)

**Wayback QA Tool:** An automated quality assurance tool for improving the quality of your captures.

**Patch Crawl:** A crawl to capture and patch in documents that were not captured in your original crawl.


**Robots.txt:** Files that a site owner can add to their site to keep crawlers from accessing all or parts of it.

## Where is my Crawl Report?

You can access your crawl reports at any time through:

- ❖ The *Crawls* link in the top navigation bar
- ❖ The *Crawls* tab within any given collection

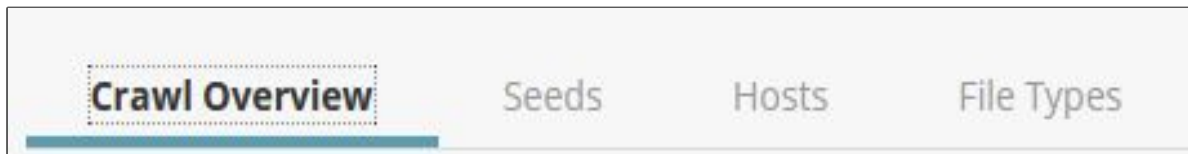
By default, crawl reports are listed by **Crawl ID**, which is a unique identifier displayed to the left of each crawl.

|         |   |
|---------|---|
| 1414202 |   |
| 1414101 |  |

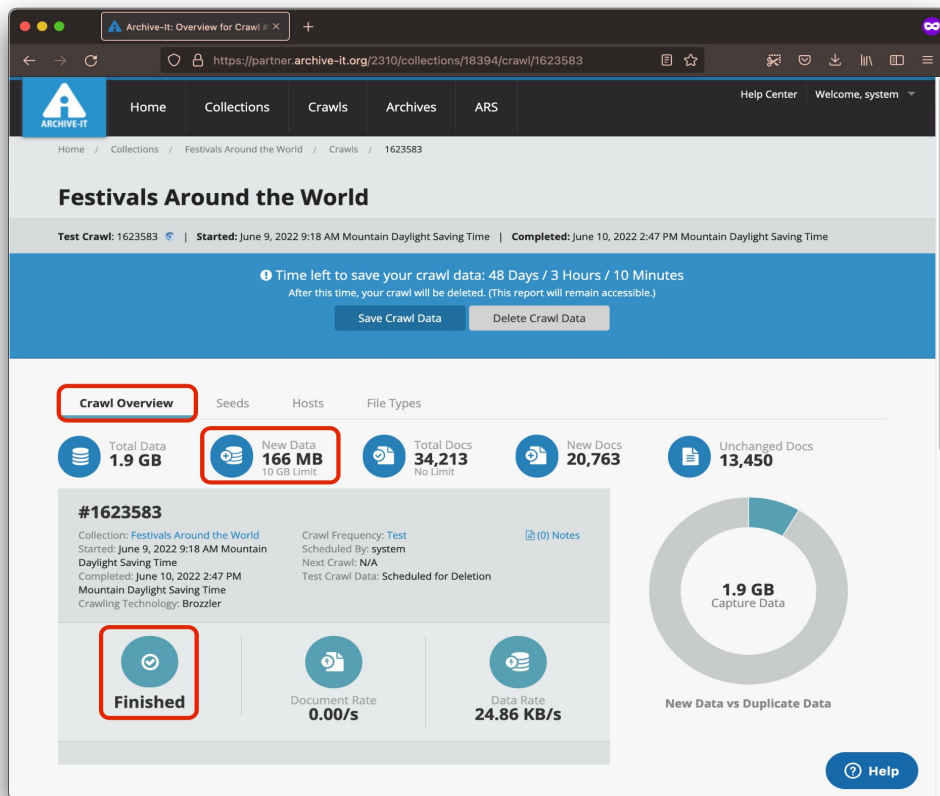


# What's inside a Crawl Report?

- ❖ Crawl Overview
- ❖ Seeds Report
- ❖ Hosts Report
- ❖ File Types Report



# REVIEW: Crawl Report Overview



Many details about the crawl, including

- *New Data*
- *Crawl Status*

Test Crawls have options to *Save or Delete Crawl Data* in the blue area

# REVIEW: Seeds Report

The screenshot shows the Archive-IT interface for a crawl titled "Festivals Around the World". The "Seeds" tab is selected and highlighted with a red box. Below the tab, there are statistics: Total Data (1.9 GB), New Data (166 MB), Total Docs (34,213), New Docs (20,763), and Unchanged Docs (13,450).

Seed List (1 Seed)

Type to Filter Seeds

[Download Seed List](#)

| Seed URL  | Seed Type | Seed Status | Docs   | New Docs | Data   | New Data | Seed                      | Wayback Link                 |
|---|-----------|-------------|--------|----------|--------|----------|---------------------------|------------------------------|
| <a href="https://twitter.com/RioCarnaval/">https://twitter.com/RioCarnaval/</a> | Standard  | Crawled     | 34,213 | 20,763   | 1.9 GB | 166 MB   | <a href="#">Seed &gt;</a> | <a href="#">Wayback &gt;</a> |

Scoping Rules for this Crawl

- [Crawl Limits](#) (2)
- [Collection Scope Rules](#) (5)
- [Seed Scope Rules](#) (3)

Seed List (1 Seed)

Type to Filter Seeds

[Download Seed List](#)

| Seed URL  | Seed Type | Seed Status | Docs   | New Docs | Data   | New Data | Seed                      | Wayback Link                 |
|---|-----------|-------------|--------|----------|--------|----------|---------------------------|------------------------------|
| <a href="https://twitter.com/RioCarnaval/">https://twitter.com/RioCarnaval/</a> | Standard  | Crawled     | 34,213 | 20,763   | 1.9 GB | 166 MB   | <a href="#">Seed &gt;</a> | <a href="#">Wayback &gt;</a> |

[Help](#)

Seeds Report list the *Seed URLs* included in the Crawl.

For each seed, it will also list

- *Seed Type*
- ***Seed Status***
- *Docs & Data* collected
- *Seed* link to Seed Management Interface
- ***Wayback*** link

Clicking directly on a *Seed URL* itself will open the Seed's Hosts Report

# REVIEW: Seeds' Hosts Report

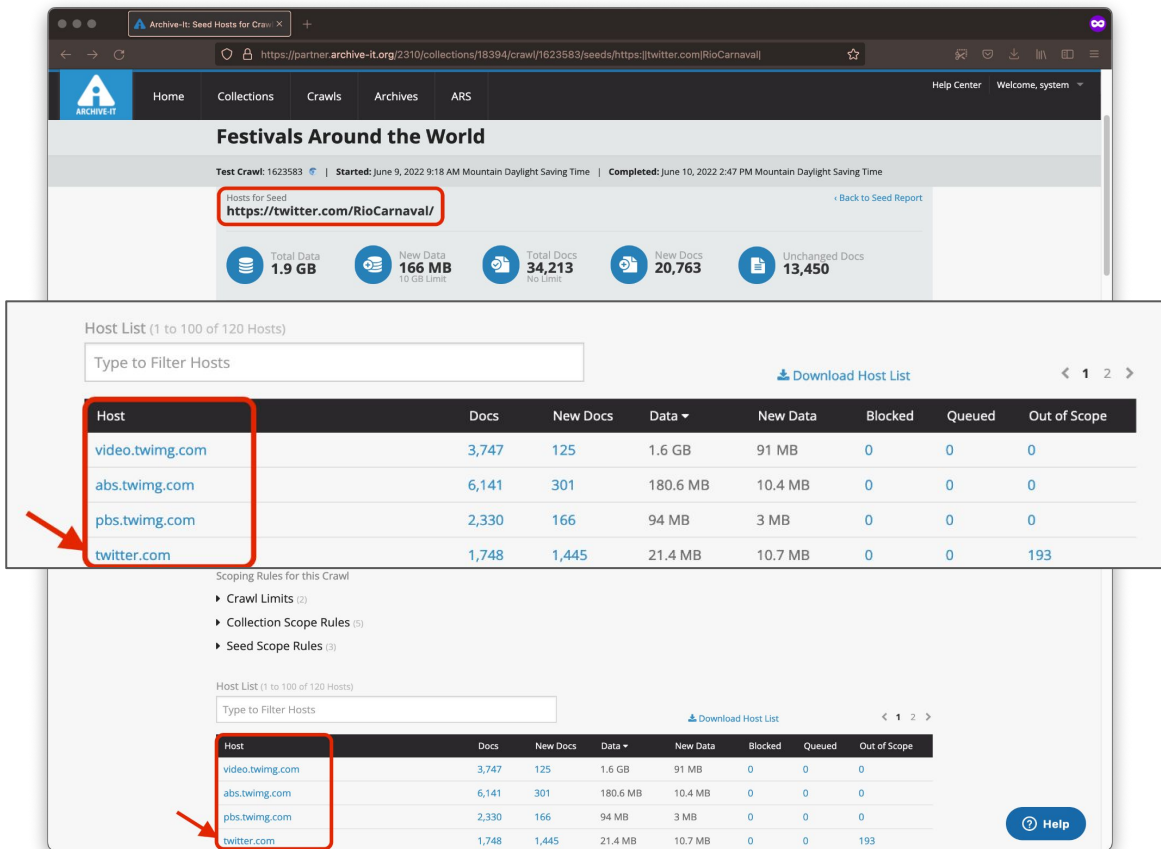
Lists each host from which content was discovered for the Seed URL on the live web

Also lists for each host:

- *Docs* collected
- *Data* totals
- *Blocked* documents
- *Queued* documents
- *Out of Scope* documents

Click directly on the *Hosts'* links to see lists of documents from a host

Click directly on numbers to see lists of documents



Archive-IT: Seed Hosts for Crawl

Home Collections Crawls Archives ARS Help Center Welcome, system

## Festivals Around the World

Test Crawl: 1623583 | Started: June 9, 2022 9:18 AM Mountain Daylight Saving Time | Completed: June 10, 2022 2:47 PM Mountain Daylight Saving Time

Hosts for Seed: <https://twitter.com/RioCarnaval/> [Back to Seed Report](#)

Total Data: 1.9 GB | New Data: 166 MB | Total Docs: 34,213 | New Docs: 20,763 | Unchanged Docs: 13,450

Host List (1 to 100 of 120 Hosts)

Type to Filter Hosts [Download Host List](#) < 1 2 >

| Host                            | Docs  | New Docs | Data     | New Data | Blocked | Queued | Out of Scope |
|---------------------------------|-------|----------|----------|----------|---------|--------|--------------|
| <a href="#">video.twimg.com</a> | 3,747 | 125      | 1.6 GB   | 91 MB    | 0       | 0      | 0            |
| <a href="#">abs.twimg.com</a>   | 6,141 | 301      | 180.6 MB | 10.4 MB  | 0       | 0      | 0            |
| <a href="#">pbs.twimg.com</a>   | 2,330 | 166      | 94 MB    | 3 MB     | 0       | 0      | 0            |
| <a href="#">twitter.com</a>     | 1,748 | 1,445    | 21.4 MB  | 10.7 MB  | 0       | 0      | 193          |

Scoping Rules for this Crawl

- ▶ [Crawl Limits](#) (2)
- ▶ [Collection Scope Rules](#) (5)
- ▶ [Seed Scope Rules](#) (3)

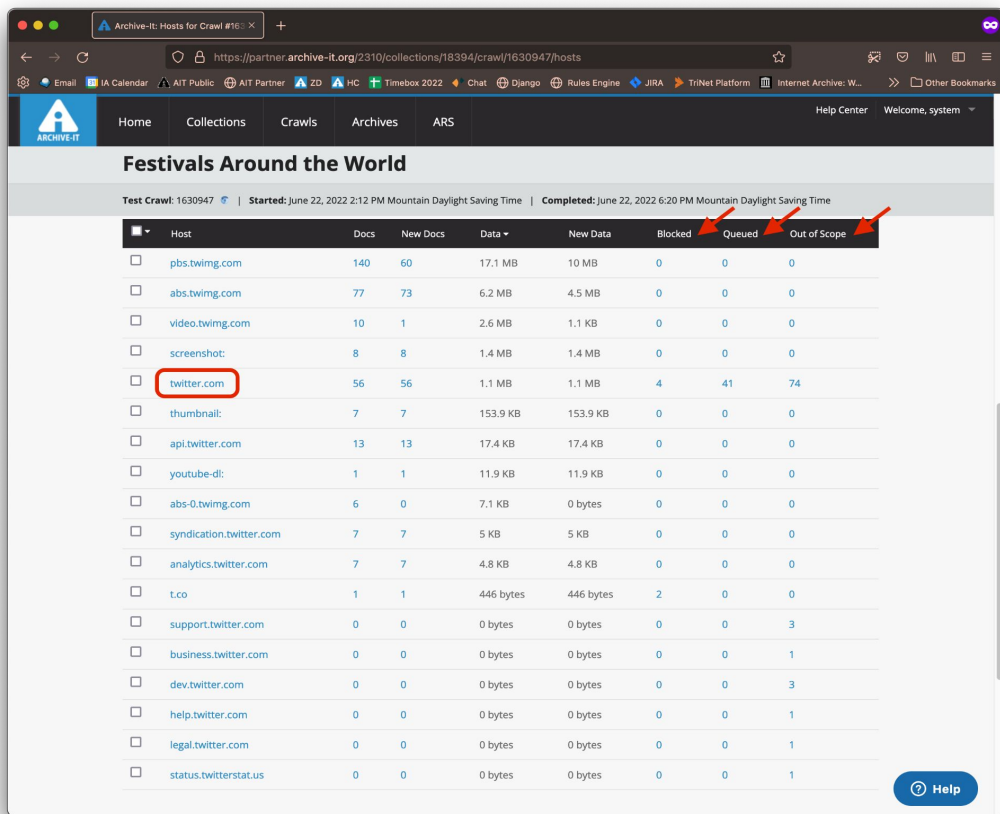
Host List (1 to 100 of 120 Hosts)

Type to Filter Hosts [Download Host List](#) < 1 2 >

| Host                            | Docs  | New Docs | Data     | New Data | Blocked | Queued | Out of Scope |
|---------------------------------|-------|----------|----------|----------|---------|--------|--------------|
| <a href="#">video.twimg.com</a> | 3,747 | 125      | 1.6 GB   | 91 MB    | 0       | 0      | 0            |
| <a href="#">abs.twimg.com</a>   | 6,141 | 301      | 180.6 MB | 10.4 MB  | 0       | 0      | 0            |
| <a href="#">pbs.twimg.com</a>   | 2,330 | 166      | 94 MB    | 3 MB     | 0       | 0      | 0            |
| <a href="#">twitter.com</a>     | 1,748 | 1,445    | 21.4 MB  | 10.7 MB  | 0       | 0      | 193          |

[Help](#)

# REVIEW: Hosts Report



Archive-IT: Hosts for Crawl #1630947

https://partner.archive-it.org/2310/collections/18394/crawl/1630947/hosts

Home Collections Crawls Archives ARS Help Center Welcome, system

### Festivals Around the World

Test Crawl: 1630947 | Started: June 22, 2022 2:12 PM Mountain Daylight Saving Time | Completed: June 22, 2022 6:20 PM Mountain Daylight Saving Time

| Host   | Docs | New Docs | Data      | New Data  | Blocked | Queued | Out of Scope |
|--|------|----------|-----------|-----------|---------|--------|--------------|
| <input type="checkbox"/> pbs.twimg.com           | 140  | 60       | 17.1 MB   | 10 MB     | 0       | 0      | 0            |
| <input type="checkbox"/> abs.twimg.com           | 77   | 73       | 6.2 MB    | 4.5 MB    | 0       | 0      | 0            |
| <input type="checkbox"/> video.twimg.com         | 10   | 1        | 2.6 MB    | 1.1 KB    | 0       | 0      | 0            |
| <input type="checkbox"/> screenshot:             | 8    | 8        | 1.4 MB    | 1.4 MB    | 0       | 0      | 0            |
| <input type="checkbox"/> <b>twitter.com</b>      | 56   | 56       | 1.1 MB    | 1.1 MB    | 4       | 41     | 74           |
| <input type="checkbox"/> thumbnail:              | 7    | 7        | 153.9 KB  | 153.9 KB  | 0       | 0      | 0            |
| <input type="checkbox"/> api.twitter.com         | 13   | 13       | 17.4 KB   | 17.4 KB   | 0       | 0      | 0            |
| <input type="checkbox"/> youtube-di:             | 1    | 1        | 11.9 KB   | 11.9 KB   | 0       | 0      | 0            |
| <input type="checkbox"/> abs-0.twimg.com         | 6    | 0        | 7.1 KB    | 0 bytes   | 0       | 0      | 0            |
| <input type="checkbox"/> syndication.twitter.com | 7    | 7        | 5 KB      | 5 KB      | 0       | 0      | 0            |
| <input type="checkbox"/> analytics.twitter.com   | 7    | 7        | 4.8 KB    | 4.8 KB    | 0       | 0      | 0            |
| <input type="checkbox"/> t.co                    | 1    | 1        | 446 bytes | 446 bytes | 2       | 0      | 0            |
| <input type="checkbox"/> support.twitter.com     | 0    | 0        | 0 bytes   | 0 bytes   | 0       | 0      | 3            |
| <input type="checkbox"/> business.twitter.com    | 0    | 0        | 0 bytes   | 0 bytes   | 0       | 0      | 1            |
| <input type="checkbox"/> dev.twitter.com         | 0    | 0        | 0 bytes   | 0 bytes   | 0       | 0      | 3            |
| <input type="checkbox"/> help.twitter.com        | 0    | 0        | 0 bytes   | 0 bytes   | 0       | 0      | 1            |
| <input type="checkbox"/> legal.twitter.com       | 0    | 0        | 0 bytes   | 0 bytes   | 0       | 0      | 1            |
| <input type="checkbox"/> status.twitterstat.us   | 0    | 0        | 0 bytes   | 0 bytes   | 0       | 0      | 1            |

Help

Hosts Reports list all hosts through which content was discovered during the crawl

- *Docs and New Docs*
- *Data and New Data*
- *Blocked Documents*
- *Queued Documents*
- *Out of Scope Documents*

# REVIEW: File Types Report

File Type List (1 to 100 of 17,316 File Types)

Type to Filter File Types

[Download List of File Types](#)

| File Type              | Docs  | New Docs | Data     | New Data |
|------------------------|-------|----------|----------|----------|
| video/mp4              | 2,372 | 79       | 1.6 GB   | 90.9 MB  |
| application/javascript | 5,960 | 357      | 181.7 MB | 12.4 MB  |
| image/jpeg             | 2,484 | 426      | 112.6 MB | 26.7 MB  |
| application/json       | 1,940 | 1,445    | 16.7 MB  | 6.4 MB   |

► Crawl Limits (2)

► Collection Scope Rules (3)

► Seed Scope Rules (3)

File Type List (1 to 100 of 17,316 File Types)

Type to Filter File Types

[Download List of File Types](#)

| File Type              | Docs  | New Docs | Data     | New Data |
|------------------------|-------|----------|----------|----------|
| video/mp4              | 2,372 | 79       | 1.6 GB   | 90.9 MB  |
| application/javascript | 5,960 | 357      | 181.7 MB | 12.4 MB  |
| image/jpeg             | 2,484 | 426      | 112.6 MB | 26.7 MB  |
| application/json       | 1,940 | 1,445    | 16.7 MB  | 6.4 MB   |

[Help](#)

The File Types report lists documents collected by the kind of files they are.

The File Type itself lists the kind of content and then the file extensions of documents, e.g. **video/mp4**

## I will show you how to:

- ❖ Find your crawl report
- ❖ Read your crawl report



## Browse Wayback pages

Check replay of your archived pages in Wayback.

Compare their look and feel with the live web's pages.

Ask yourself:

- ❖ Does it meet expectations?
- ❖ Will it serve future users' needs?

Try some QA strategies if not!



## Browse Wayback pages

Look for the same features as the Live Web's counterpart website:

❖ **Styling features**

headers/footers  
grids/boxes

.css

❖ **Interactive features**

dropdown menus  
carousels/sliders

.js (javascript)

❖ **Embedded elements**

images  
videos  
audio files

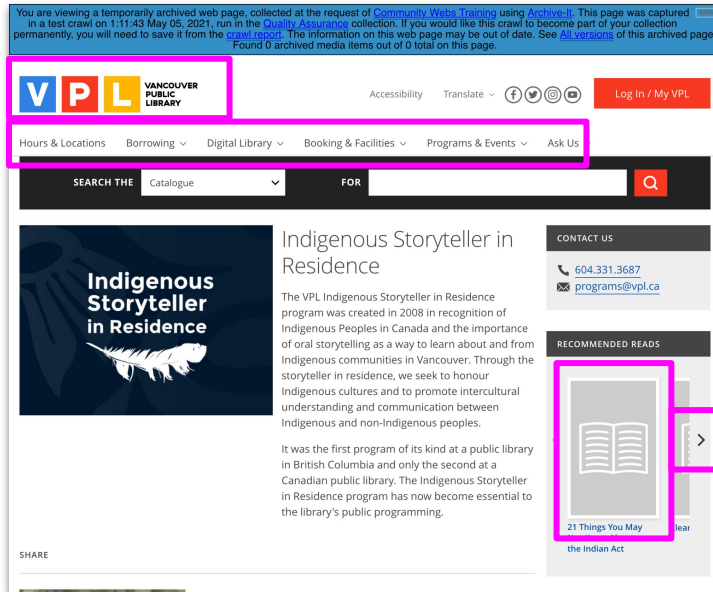
.jpg, .jpeg, .png  
.mp4, .mp2t, .mpeg (video)  
.mp3, .mpeg (audio)

❖ **Working links**

.html, .js

# Browse Wayback pages

Look for the same features



as the Live Web's counterpart



A patch crawl is a crawl to capture and patch in documents that were not captured in your original crawl.

Two ways to run patch crawls:

- ❖ Via the Hosts Report
- ❖ Via the Wayback QA tool

# PATCH CRAWL via the Hosts Report

## Hosts Report's Blocked Documents

Archive-IT: Hosts for Crawl #1: x

partner.archive-it.org/2021/collections/16622/crawl/1412828/hosts

Home Collections Crawls Archives ARS

QA May 13

One-Time Crawl: 1412828 | Started: May 7, 2021 5:31 PM Mountain Daylight Time | Completed: May 8, 2021 5:33 AM Mountain Daylight Time

Host List (1 to 100 of 633 Hosts)

Type to Filter Hosts

Edit Rules Run Patch Crawl

| Host                            | Docs  | New Docs | Data      | New Data  | Blocked | Quoted | Out of Scope |
|---------------------------------|-------|----------|-----------|-----------|---------|--------|--------------|
| www.upi.ca                      | 2,902 | 1,963    | 288.8 MB  | 230.5 MB  | 103     | 1,817  | 0            |
| rs--sn-nlv7nvt7.googlevideo.com | 2     | 1        | 57.3 KB   | 57.3 KB   | 0       | 0      | 7            |
| rs--sn-a5melnzr.googlevideo.com | 2     | 1        | 15.5 MB   | 15.5 MB   | 0       | 0      | 1            |
| code.jquery.com                 | 8     | 4        | 1.6 MB    | 832.2 KB  | 0       | 0      | 14           |
| m.youtube.com                   | 4     | 4        | 562.2 KB  | 562.2 KB  | 0       | 0      | 19           |
| s7.addthis.com                  | 4     | 3        | 354.4 KB  | 354 KB    | 0       | 0      | 7            |
| www.youtube.com                 | 7     | 7        | 161.7 KB  | 161.7 KB  | 0       | 0      | 90           |
| www.googletagmanager.com        | 4     | 4        | 113.8 KB  | 113.8 KB  | 0       | 0      | 5            |
| services.angononline.com        | 4     | 1        | 43.9 KB   | 21.5 KB   | 0       | 0      | 2            |
| www.enr.com                     | 4     | 4        | 40.5 KB   | 40.5 KB   | 0       | 0      | 59           |
| www.google.com                  | 3     | 2        | 10.8 KB   | 3.2 KB    | 0       | 0      | 15           |
| www.google.ca                   | 1     | 1        | 7.6 KB    | 7.6 KB    | 21      | 0      | 2            |
| wholc                           | 2     | 2        | 3.9 KB    | 3.9 KB    | 0       | 0      | 32           |
| rs--sn-nlv7nvt7.googlevideo.com | 2     | 1        | 3 KB      | 1.2 KB    | 0       | 0      | 3            |
| forms.googleapps.com            | 4     | 3        | 2.6 KB    | 2.3 KB    | 0       | 0      | 5            |
| script.crazyegg.com             | 5     | 1        | 2.4 KB    | 526 bytes | 0       | 0      | 2            |
| dnsc                            | 24    | 24       | 2.3 KB    | 2.3 KB    | 0       | 0      | 0            |
| use.typekit.net                 | 2     | 2        | 1.8 KB    | 1.8 KB    | 0       | 0      | 5            |
| p.typekit.net                   | 2     | 2        | 646 bytes | 646 bytes | 0       | 0      | 1            |

Help

## Start a Patch Crawl for Blocked Documents

Archive-IT: Hosts for Crawl #1: x

partner.archive-it.org/2021/collections/16622/crawl/1412828/hosts

Home Collections Crawls Archives ARS

QA May 13

One-Time Crawl: 1412828 | Started: May 7, 2021 5:31 PM Mountain Daylight Time | Completed: May 8, 2021 5:33 AM Mountain Daylight Time

Host List (1 to 100 of 633 Hosts)

Type to Filter Hosts

Edit Rules Run Patch Crawl

Run Patch Crawl (1 Host Selected)

This patch crawl will capture URLs from specific hosts that were blocked by robots.txt, if the Ignore Robots.txt check box is selected. Once started, the patch crawl can be monitored from the Current Crawls area of your account. Once it has finished, a full set of reports will be generated.

☒ Ignore Robots.txt

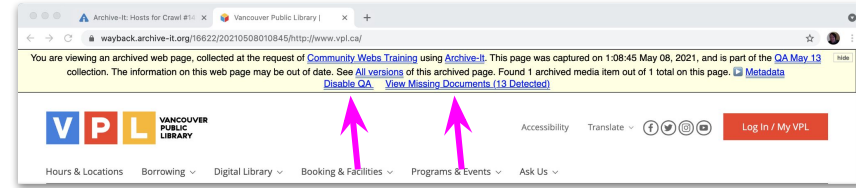
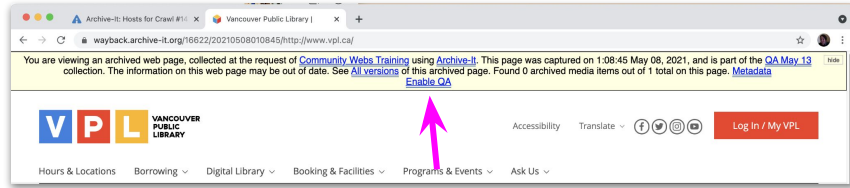
Cancel Run Patch Crawl

Filter Selected Hosts by Keyword

| Host                            | Docs  | New Docs | Data      | New Data  | Blocked | Quoted | Out of Scope | Remove |
|---------------------------------|-------|----------|-----------|-----------|---------|--------|--------------|--------|
| www.upi.ca                      | 2,902 | 1,963    | 288.8 MB  | 230.5 MB  | 103     | 1,817  | 0            |        |
| m.youtube.com                   | 4     | 4        | 562.2 KB  | 562.2 KB  | 0       | 0      | 19           |        |
| s7.addthis.com                  | 4     | 3        | 354.4 KB  | 354 KB    | 0       | 0      | 7            |        |
| www.youtube.com                 | 7     | 7        | 161.7 KB  | 161.7 KB  | 0       | 0      | 90           |        |
| www.googletagmanager.com        | 4     | 4        | 113.8 KB  | 113.8 KB  | 0       | 0      | 5            |        |
| services.angononline.com        | 4     | 1        | 43.9 KB   | 21.5 KB   | 0       | 0      | 2            |        |
| www.enr.com                     | 4     | 4        | 40.5 KB   | 40.5 KB   | 0       | 0      | 59           |        |
| www.google.com                  | 3     | 2        | 10.8 KB   | 3.2 KB    | 0       | 0      | 15           |        |
| www.google.ca                   | 1     | 1        | 7.6 KB    | 7.6 KB    | 21      | 0      | 2            |        |
| wholc                           | 2     | 2        | 3.9 KB    | 3.9 KB    | 0       | 0      | 32           |        |
| rs--sn-nlv7nvt7.googlevideo.com | 2     | 1        | 3 KB      | 1.2 KB    | 0       | 0      | 3            |        |
| forms.googleapps.com            | 4     | 3        | 2.6 KB    | 2.3 KB    | 0       | 0      | 5            |        |
| script.crazyegg.com             | 5     | 1        | 2.4 KB    | 526 bytes | 0       | 0      | 2            |        |
| dnsc                            | 24    | 24       | 2.3 KB    | 2.3 KB    | 0       | 0      | 0            |        |
| use.typekit.net                 | 2     | 2        | 1.8 KB    | 1.8 KB    | 0       | 0      | 5            |        |
| p.typekit.net                   | 2     | 2        | 646 bytes | 646 bytes | 0       | 0      | 1            |        |

Help

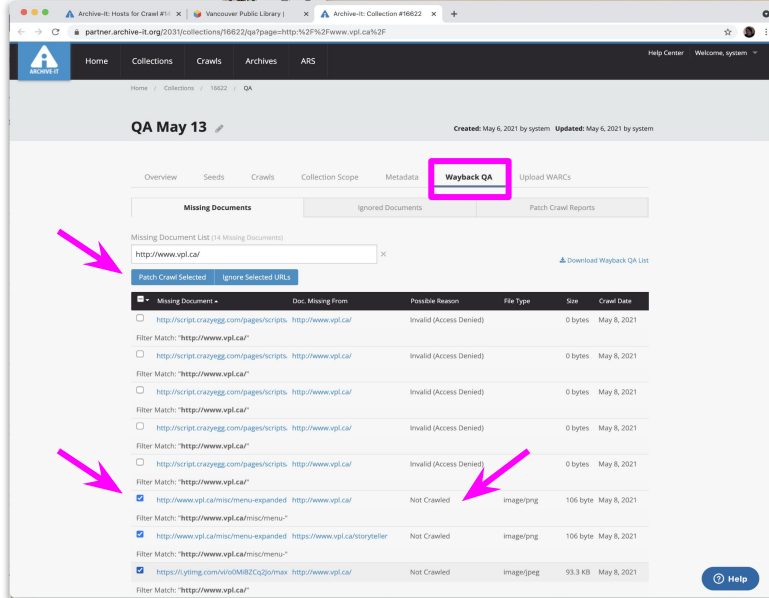
## The Wayback QA tool



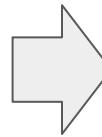
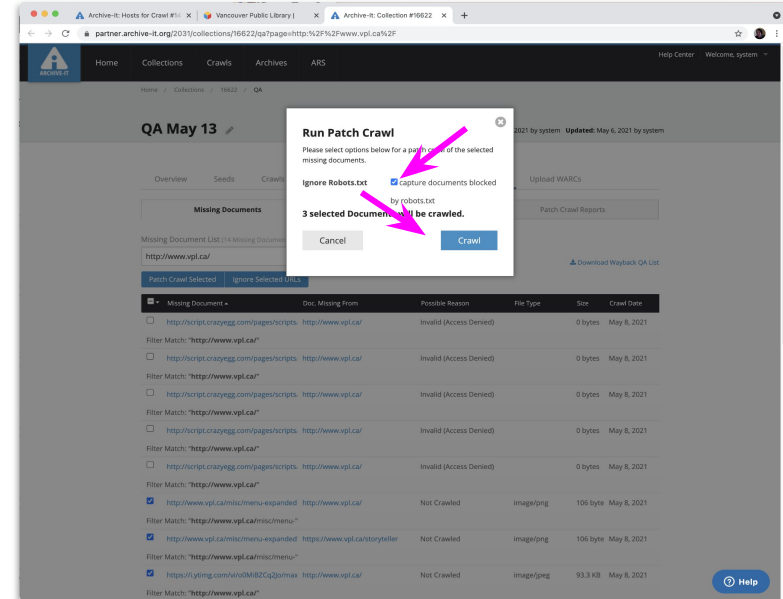
- ❖ Scans for missing documents on the Wayback page
- ❖ Allows patch crawls for missing documents
- ❖ Used for Production Captures (Not Test Captures)

# Patch Crawl with Wayback QA

In the partner account QA application



The screenshot shows the 'Wayback QA' application interface. The 'Wayback QA' tab is highlighted with a pink box. Below the 'Missing Documents' section, there is a table of missing documents. The table has columns: Doc. Missing From, Possible Reason, File Type, Size, and Crawl Date. The first row shows a document from 'http://www.vpl.ca/' with a 'Possible Reason' of 'Invalid (Access Denied)'. The second row shows a document from 'http://script.craeygg.com/pages/scripts/' with a 'Possible Reason' of 'Invalid (Access Denied)'. The third row shows a document from 'http://www.vpl.ca/misc/menu-expanded' with a 'Possible Reason' of 'Not Crawled'. The fourth row shows a document from 'http://www.vpl.ca/misc/menu-' with a 'Possible Reason' of 'Not Crawled'. The fifth row shows a document from 'http://www.vpl.ca/misc/menu-' with a 'Possible Reason' of 'Not Crawled'. The sixth row shows a document from 'https://i.yimg.com/nvOM8ZCq2oImax' with a 'Possible Reason' of 'Not Crawled'. The table is filtered by 'Filter Match: "http://www.vpl.ca/"'. There are pink arrows pointing to the 'Wayback QA' tab, the 'Missing Documents' section, and the 'Not Crawled' status in the table.

The screenshot shows the 'Wayback QA' application interface with the 'Run Patch Crawl' dialog box open. The dialog box has a title 'Run Patch Crawl' and a message 'Please select options below for a patch crawl of the selected missing documents.' There are two checkboxes: 'Ignore Robots.txt' and 'Capture documents blocked by robots.txt'. The 'Capture documents blocked by robots.txt' checkbox is checked. Below the checkboxes, it says '3 selected Documents will be crawled.' There are 'Cancel' and 'Crawl' buttons. There are pink arrows pointing to the 'Run Patch Crawl' dialog box, the 'Capture documents blocked by robots.txt' checkbox, and the 'Crawl' button.

# QUALITY ASSURANCE

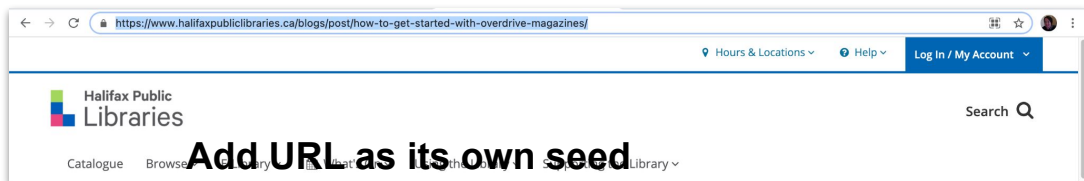
## Use the Wayback QA tool

Detects  
missing  
documents  
on the page  
and allows  
you to patch  
crawl them

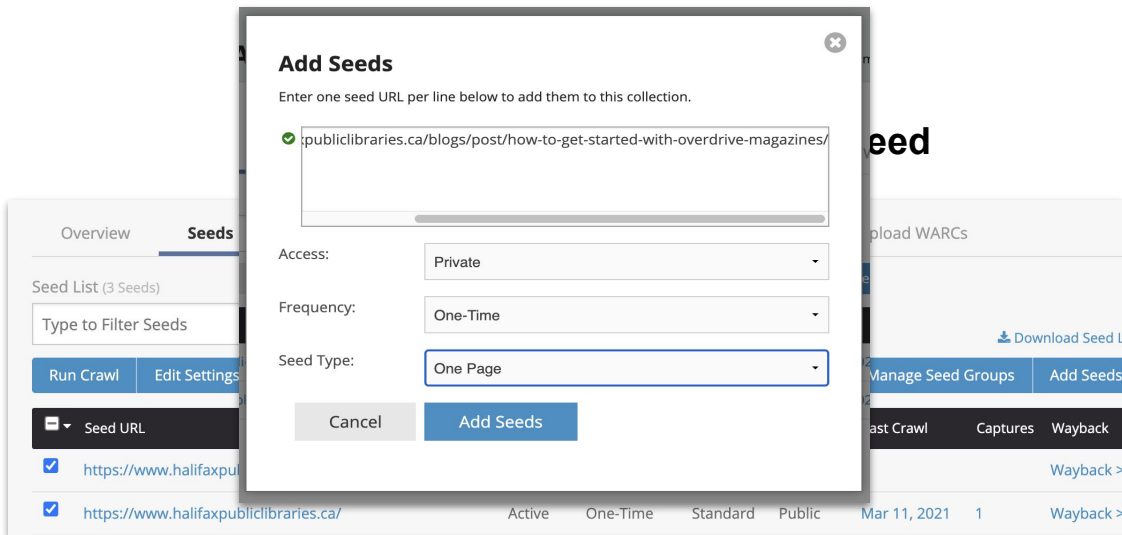
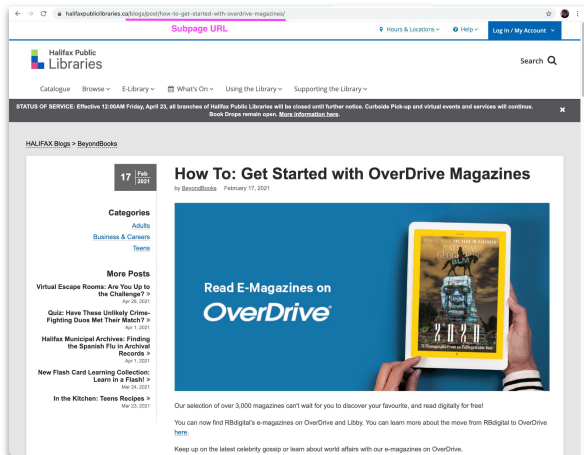
# HELPER SEEDS

<https://www.halifaxpubliclibraries.ca/blogs/post/how-to-get-started-with-overdrive-magazines/>

## Copy subpage URL



## Add URL as its own seed





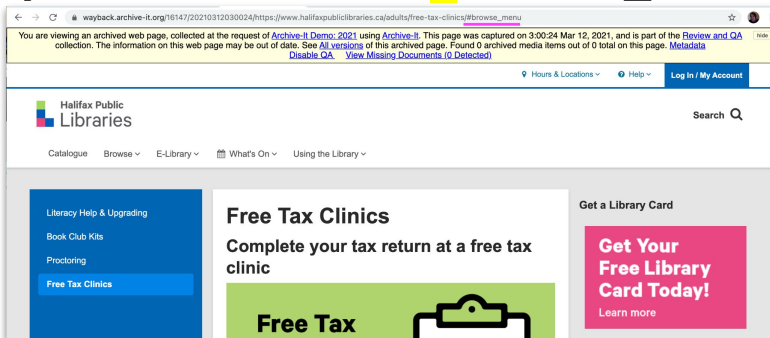
# Open Wayback pages in NEW TABS



Useful for opening

- ❑ Video Watchpages from a YouTube Channel
- ❑ Links built with Javascript
- ❑ Subpage URLs with a #:

[https://www.halifaxpubliclibraries.ca/#browse\\_menu](https://www.halifaxpubliclibraries.ca/#browse_menu)



## I will show you how to:

- ❖ Load and look over the Wayback page
- ❖ Use the Wayback QA Tool to patch crawl
- ❖ Patch Crawl via the Hosts Report



# QUALITY ASSURANCE

## Leverage the Help Center

Review the documentation in our Help Center anytime:

### [Scoping guidance for specific kinds of sites](#)

- What is the usual scoping?
- What is the expected replay?

### [System Status/Social media and other platforms status](#)

- How are systems and platforms currently performing?

### [Known Web Archiving Challenges](#)

- Is this content archivable?

### [Troubleshooting Browser Issues](#)

- Is my browser interfering with replay?

### [What to do when you see a blank page in Wayback](#)

- Have I tried Brozzler?
- Is JavaScript interfering?

## QA CHECKLIST

- ☐ Prioritize: Audiovisual & dynamic content; significant properties
- ☐ Check crawl and seed status, data, and docs.
- ☐ Replay and browse with Wayback.
- ☐ Find missing documents: blocked, queued, and “out of scope.”
- ☐ Patch crawl missing documents for prioritized pages.
- ☐ Leverage the Help Center, particularly System Status page.
- ☐ When in need, ask for help 🙋

## LEARN MORE





Photo courtesy of Towfiqu Barbhuiya on Unsplash

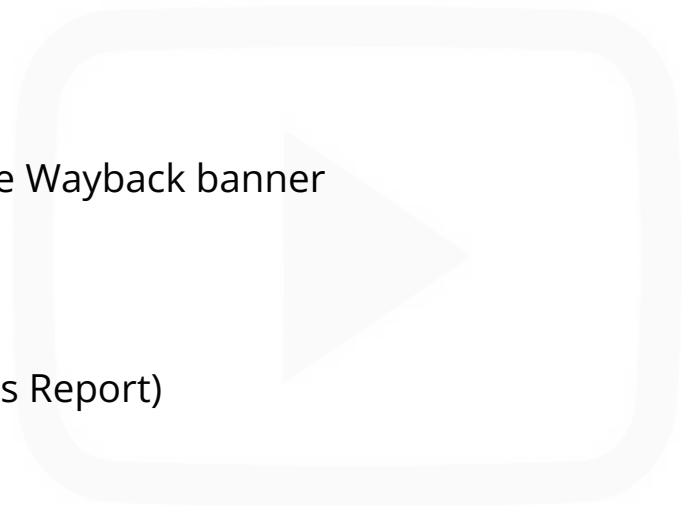
Help Center: <https://support.archive-it.org/hc/en-us>

Check out our blog: [www.archive-it.org/blog](http://www.archive-it.org/blog)

**Thank you! Any questions?**



- **Watch pages / embedded videos** 
- **Channels / users/ playlists** 
  - Videos may need to be opened in new browser tabs in order to replay
  - Videos from playlists may not replay in Wayback
- **Expectations**
  - Replays most reliably in Chrome
  - Videos will replay in-page and via “Videos” link in the Wayback banner
- **Troubleshooting**
  - Did I use Brozzler?
  - [yt-dlp](#): did it run? (look for host “youtube-dl” in Hosts Report)
  - File Types Report: were video files collected?





- open-source command-line utility for retrieving media which enables the preservation of full, discrete files (ie. MP4, MP3, MKV, etc.) for access
- Runs on web pages during the crawling process and deposits media items it finds into WARC files with corresponding JSON metadata
- When it runs as expected:
  - Wayback can load information from its corresponding JSON metadata files into the banner message
  - specify how many media items were archived
  - provide direct access to media via a media player

# youtube-dl

You are viewing an archived web page, collected at the request of [Karl-Rainer Blumenthal](#) using [Archive-It](#). This page was captured on 16:46:34 Feb 04, 2020, and is part of the [Demo](#) collection. The information on this web page may be out of date. See [All versions](#) of this archived page.

hide

Found 2 archived media items out of 2 total on this page. [▶ Metadata](#)

[Enable QA](#)

| <input type="checkbox"/> | Host  | Docs ▲ | New Docs | Data      | New Data  |
|--------------------------|---|--------|----------|-----------|-----------|
| <input type="checkbox"/> | <a href="#">support.google.com</a>                | 0      | 0        | 0 bytes   | 0 bytes   |
| <input type="checkbox"/> | <a href="#">rr3---sn-o097znss.googlevideo.com</a> | 1      | 1        | 1.2 KB    | 1.2 KB    |
| <input type="checkbox"/> | <a href="#">r1---sn-h5qzen7y.googlevideo.com</a>  | 1      | 1        | 5.5 KB    | 5.5 KB    |
| <input type="checkbox"/> | <a href="#">rr3---sn-o097znze.googlevideo.com</a> | 1      | 1        | 193 MB    | 193 MB    |
| <input type="checkbox"/> | <a href="#">i1.ytimg.com</a>                      | 2      | 2        | 85.5 KB   | 85.5 KB   |
| <input type="checkbox"/> | <a href="#">r1---sn-o097znze.googlevideo.com</a>  | 2      | 2        | 646 bytes | 646 bytes |
| <input type="checkbox"/> | youtube-dl:                                       | 3      | 3        | 1.1 MB    | 1.1 MB    |
| <input type="checkbox"/> | <a href="#">accounts.google.com</a>               | 3      | 3        | 1.5 MB    | 1.5 MB    |
| <input type="checkbox"/> | <a href="#">static.doubleclick.net</a>            | 3      | 3        | 2.5 KB    | 2.5 KB    |
| <input type="checkbox"/> | <a href="#">youtube.com</a>                       | 3      | 3        | 3.8 KB    | 3.8 KB    |
| <input type="checkbox"/> | <a href="#">play.google.com</a>                   | 3      | 3        | 2.6 KB    | 2.6 KB    |



## Additional Resource Roundup

- [Archive-It System and Platform Status page](#)
- [Help Center articles for specific platforms](#)
- [What to do when you see a blank page](#)
- [Guide to A/V web archiving with youtube-dl](#)
- [How to use the Wayback QA tool](#)
- [How to clear your cache, cookies, and history](#)
- [Upload WARCs](#)
- [The Wayback Machine](#)